

Title Page

External examining: fit for purpose?

Sue Bloxham*, University of Cumbria, Lancaster, LA1 3JD, Tel +44 (0)1524

384384, E-mail susan.bloxham@cumbria.ac.uk

Margaret Price, Oxford Brookes University, Wheatley Campus, Oxford, OX33

1HX, Tel: +44 (0)1865 48 58 58, E-mail: meprice@brookes.ac.uk

***Corresponding author**

Abstract

In a context of international concern about academic standards, the practice of external examining is widely admired for its role in defending standards. Yet a contradiction exists between this faith in examining and continuing concerns about standards. This article argues that external examining rests on assumptions about standards which are significantly open to challenge. Six assumptions relating to the conceptual context, the operation and the nature of examiners themselves are analysed drawing on a review of the available evidence. The analysis challenges the notion of a consensus on standards and the potential to vest in individuals the ability to represent that consensus when judging the comparability of academic standards in a stable and appropriate way. The issues raised have relevance to the UK and to other national systems using external examiners or seeking to guarantee academic standards by, in some cases, adopting quality assurance approaches developed in the UK.

External examining: fit for purpose?

External examining is a process by which the fairness of assessment and the academic standards in higher education are monitored. This monitoring normally involves an academic from another institution approving assessment tasks and examinations and checking a sample of student work for the courses which they examine. A wider range of activities may also be included as discussed below. In certain limited circumstances, for example in viva voces for research degrees, external examiners directly examine student performance. In the case of vocational courses, some examiners may be practising members of the relevant profession rather than academics. External examining is prevalent in the UK and several other countries operate or have operated an external examiner system: Ireland, New Zealand, Denmark, India, Malaysia, Brunei, Malawi, Hong Kong and South Africa (Harvey 2004-2012), although examiners' roles differ across systems (Ross 2009, 475). Whilst this article necessarily focuses on the UK, the issues have wider relevance to other systems using external examiners and to other nations seeking to guarantee academic standards by, in some cases, adopting quality assurance approaches developed in the UK.

It is claimed that British external examiner system provides 'one of the principal means for maintaining UK threshold academic standards within autonomous higher education institutions' (QAA 2012: B7, 2) through an accountability process which specifically addresses the quality of student performance. Indeed, Gaunt (1999) suggests that external examining is the only quality assurance mechanism that directly addresses quality of student output and argues that it is the most highly regarded of UK accountability measures. However while the system is widely admired (Watson 2006) for its role in defending academic standards there are increasing concerns about those standards in the UK and elsewhere (HEFCE 2009; Innovation, Universities, Science and Skills Committee 2009; QAA 2009; Coates 2010). This apparent contradiction warrants closer scrutiny and

this article argues that the relationship between external examining and academic standards rests on assumptions which are significantly open to challenge.

The language of the new UK Quality Code for higher education (QAA 2012) appears to signal a tentative down grading of the external examiner as the prime guardian of standards and a subtle transformation in official views of the role. The rationale for changing the scope of external examiners' responsibilities may be an attempt to tackle continuing criticisms of the operation of the system, echoing as it does the recommendations of the most recent enquiry into the role by Universities UK (2011). However, the Universities UK report and the new code that has emerged remain silent on crucial assumptions about the nature of academic standards as understood and applied by external examiners. So, at a time when the higher education system is changing to include more private providers and will be populated by fee paying students with high expectations of reliability and fairness in marking, this is a timely point to critically examine those assumptions that underpin the system.

Whilst there have been various studies of external examining, those that consider examiners' consistency in both understanding and applying academic standards have been singularly lacking in recent years with the exception of Colley and Silver (2005). The last significant academic enquiry dates back to Silver et al. (1995) and Warren Piper (1994). On the other hand, official reports and commentaries which discuss external examining abound. Despite this plethora of enquiries and recommendations, reforms have been criticised for their limited impact in tackling the failings of the system (Attwood 2010). The aim of this article is to test the premise for perceiving external examiners as a mainstay for protecting academic standards by examining the assumptions on which the system is based. It differs from recent previous enquiries on external examining by focusing on academic standards rather than examining procedures. It also differs from official reports in adopting a scholarly approach; applying developing educational research in the field of higher education assessment and marking to the analysis.

Purpose and history of external examining

External examining is a mandatory, 'long standing and crucial feature of British Higher Education' (HEQC 1996a, 7). It has existed in some form since the early nineteenth century, gradually becoming adopted by new universities as a means of demonstrating the credibility of their standards (Silver and Williams 1996). However in recent decades a lack of confidence in academic standards has emerged, largely prompted by massification of higher education, and has led to concerns that the existing system is no longer able to warrant comparable standards across universities (Silver and Williams 1996). It is a system that is seen to be 'under strain' (HEFCE, 2009) and in need of 'repair' (Innovation, Universities, Science and Skills Committee 2009). Criticisms have tended to focus on a lack of consistency in examiners' appointment and role (HEQC 1994) and the variability in examining practices in different programmes, subject disciplines and universities, resulting from weak or inconsistent institutional processes (QAA 2005). In addition they express worries about the potential for 'cosy' relationships between examiners and departments (Universities UK and others 2010, paras 20 and 36). There are also concerns about clarity (HEFCE 2009) and authority (BIS 2009) in the examiners' role in assuring standards (HEFCE 2009) as they move to more of a critical friend role than one of arbiter of standards.

Some reports debate the notion of national standards. In the 1990s, Silver and Williams (1996, 42) concluded that there was little acceptance of the idea of 'national' standards, but they considered that external examining was more feasible as a means of comparing standards at the level of the subject. Examiners' appointment from similar types of institutions, or with similar programmes and assessment methods means that 'comparability of standards can only be achieved within fairly narrow and clearly defined parameters' (HEQC 1996a, para 24.2). But, even with that proviso,

Gaunt (1999) points to the difficulty of external examiners providing for reliability. As Sadler (2011) points out, calibration' of individuals' assessment standards in general is lacking.

The recent Universities UK (2011) report on external examiners seems to acknowledge this when it states that examiners cannot guarantee strict comparability of standards between programmes, only broad comparability with an emphasis on programme diversity but consensus over 'threshold' standards (para 12). This is reflected in the new Quality Code for UK higher education which states that external examiners:

'are also able to offer an informed view of how standards compare with the same or similar awards at other higher education institutions (primarily in the UK, and sometimes overseas as well) of *which they have experience.*' (QAA UK Quality Code for higher education, chapter B7 external examining, p2 – our emphasis)

A further indication that the role of the examiner is now considered to be less authoritative figures elsewhere in the new UK Quality Code for UK higher education (QAA 2012) which states:

'the external examiner is only one of a number of sources of external opinion which institutions consider, including external advice received through approval, monitoring and review processes. External examiners are therefore not the sole guarantors of the comparability of standards achieved.' (p9)

However, whilst the contribution of external examiners has arguably been reduced, none of these successive reports or research projects rejects the external examiner model. Furthermore, they do not present a fundamental challenge to the notion that a system of independent examiners can make a major contribution to safeguarding standards, albeit within a more limited domain. For example, Universities UK (2011) states 'The overwhelming body of opinion from the responses is

that the external examining system works well on the whole, and does not require a major overhaul' (para 8). Instead, each report has recommended improvements in the system. For example the Dearing report (1997) recommended a pool of accredited examiners, consistent training for external examiners and, archiving scripts to help maintain standards over time. More recently, the QAA (2009) recommended that the External examiner system should be more transparent and better explained; there should be improved procedures to identify and support external examiners and a set of minimum expectations of those holding the role. The Universities UK report in 2011 recommends greater transparency, minimum expectations and criteria for the appointment of examiners, and enhanced induction. In addition, it recommends core content in examiners' reports and a right of access to those reports by students on the programme. In general, diagnoses of external examiner problems have been repetitive and the solutions, we would argue, have concentrated on fixing the easier challenges of protocols and practice (Universities UK 2010; 2011). Such reports and inquiries fail to investigate the more fundamental question of whether examiners understand and can consistently apply academic standards in the way required by their, albeit possibly confused, role as defenders of academic standards. This is even though, as Brooks (2012) asserts, most concerns about standards can be traced back to the judgement processes of markers and examiners.

This policy context forms the background for the remainder of this paper in which we shall consider this fundamental question about the relationship between external examining and academic standards by considering the assumptions that underpin the system. We shall do this in three domains: the conceptual context for examining, the operation of the system and the nature of examiners, themselves. The assumptions selected have been judged by the authors to be implicit in the policy and its operation and the justification for choosing each one is outlined in the relevant section below.

The conceptual context

The notion of an external examiner system traditionally rests on a conceptualisation of standards as something about which there can be a community consensus. Such a community consensus is a necessary pre-requisite to individuals making judgements about appropriate academic standards and the comparability of outputs by students across institutions.

Assumption 1

A community consensus on academic standards can have enduring stability over time and across programmes, departments and universities;

In order to discuss this assumption, it is necessary to consider what we mean by academic standards. For the purposes of this article, we use the term 'academic standards' to describe standards of student attainment as this 'output' definition enjoys a broad consensus (Bloxham and Boyd 2011). This definition also distinguishes academic standards from quality standards, the latter of which apply to aspects such the qualifications of staff, the design of the curriculum or the availability of resources as indicators of the quality of a university programme . However, questions about the meaning of academic standards remain. Indeed, as an HEQC report (1996c, 1) states 'The absence of a clear understanding of the meaning of academic standards in the expanded, diversified and modularised world of higher education' hinders the debate about comparability.

Customary descriptions of external examining hold conceptions of standards which lie in a technorationalist standards discourse. They are based on assumptions that 'knowledge is monolithic, static and universal' (Delandshere 2001, 127) with staff acculturated into 'guild knowledge' (Sadler 1989).

On the basis of this epistemology, although statements of standards were rarely articulated in the past, there was an assumption pre-massification (Silver and Williams 1996, 27) that a 'gold

standard' existed; fixed benchmarks which enshrined the standards of the ancient universities. Such a perspective sees standards as independent of the individuals who created or are custodians of them. Whilst the notion of national standards has since been disputed, the underpinning discourse of quality assurance and examining continues to privilege this techno-rationalist conception. Yorke (2008, 27) links this discourse to a 'measurement' model and a 'realist' perspective on assessment.

However, such discourse sits in contrast to the hermeneutic approach which is 'grounded in the philosophical and scientific tradition of phenomenology' (Broad 2000, 244). This tradition includes a socio-cultural perspective which 'emphasises the situatedness of practice' and a social constructivist approach which 'emphasises the constructedness of knowledge' (Shay 2008, 596). Such a philosophy is interpretive and conceives of staff as active agents in co-creating standards through local assessment practices and communities. It argues that, in contrast to the gold standard approach, there is growing theoretical argument and empirical evidence for the notion that individuals construct their own 'standards frameworks' (Ashworth et al. 2010). This occurs despite the evidence that tutors think their standards are the same (Warren Piper 1994, 79). This conceptualisation of 'standards frameworks' is supported by work in the broader fields of higher education research on assessment and marking. For example, similar notions of this concept have been generated such as '[t]eachers' conceptions of quality' (Sadler 1989), assessors' 'evaluative frameworks' (Broad 2003), 'assessors' interpretive frameworks' (Shay 2005, 665) individuals' 'concepts and interpretations' (Wolf 1995, 67-8) and 'pre-formed knowledge structures' (Crisp 2008, 250). Such perspectives emerge from different traditions but they all emphasise the interpretive, personalised and subjective nature of standards, learnt informally through active participation in relevant communities and practices and lacking the characteristics that would enable a broad and enduring consensus. Yorke (2008) links this paradigm to a 'judgement' model and a 'relativist' perspective on assessment.

As Shay (2005) argues, the interpretive framework that each examiner holds combines objective elements learnt from participation in the relevant field, and highly subjective elements arising from the local context of assessment and dependent on the assessor. 'From this perspective assessment is situationally contingent, rooted in local cultures and reliable and robust only in terms of sets of assumptions, attitudes and values which are, in part at least, localized' (2005, 669). Such standards are not 'immutable' (Sadler 1987, 196) or independent of individuals (Shay 2005, 677). These individualised standards appear to survive regardless of social opportunities to negotiate and construct shared standards. For example, numerous studies over time show low agreement between higher education assessors (Sadler 1987; Elander and Hardman 2002; Bloxham 2009; Shay 2004; Leach et al. 2001; Wolf 1995). As Broad (2003, 74) states, research shows that 'standards refuse[d] to be as solid, stable and portable an entity as participants wished'.

Fundamentally, the notion that we can establish standards independently from the individuals who are using them is challenged:

'We are social beings who construe the world according to our values and perceptions; thus, our biographies are central to what we see and how we interpret it. Similarly in assessment, performance is not "objective"; rather, it is construed according to the perspectives and values of the assessor'. (Gipps 1999, 370)

Thus, from this latter perspective, the simplistic and fixed notions of standards as portrayed in public debate deny the necessarily elusive and dynamic nature of academic standards which are continuously co-constructed by academics.

Whether one adopts a techno-rational or hermeneutic approach to standards, the issue of consensus is further hampered by the fact that standards generally constitute tacit as well as explicit knowledge; their oral and implicit nature meaning that it is difficult to detect drifts in standards (Sadler 1987, 192). Therefore, although at a fairly general and abstract level, higher education

assessors do share a significant measure of agreement about what they are looking for in student work (Warren Piper 1994), that does not establish consensus regarding the standards by which these aspects of work will be assessed. Despite the view of Silver and Williams (1996, 42) that a 'more attainable goal' is a broad view of external examining with more 'vague and indefinable' standards (p44), reports on standards in external examining continue to presume a consensus on academic standards, at least at a discipline/ programme level. However the new UK Quality Code (QAA 2012) indicates that examiners' contribution should be to the management of 'threshold' standards (p4). The insertion of 'threshold' in relation to standards is interesting in itself in what it signals about trust in examiners' capacity to reflect community consensus regarding standards although why 'threshold' standards are more important or easier to determine than, for example differences between grades, is not made clear.

The operation

In this section, we will consider two assumptions that underpin the effective operation of the external examiner system.

Assumption 2

There is a common understanding of the external examiner role.

In recent years, the role of external examiner has become more diverse and includes conceptions of the examiner as arbiter of standards, external reference point and critical friend. In 1996, the HEQC (1996a) study found four conceptions of the examiners' role: additional examiner, moderator, calibrator and consultant and the Silver report (Silver et al. 1995) found mixed views about extending the role towards a more advisory or consultancy function for both principled and

pragmatic reasons. In more recent times, as Yorke (2008, 103) notes, examiners have become 'more remote' from students' actual work as programmes of study have become modularised. Their role has shifted away from looking at samples of students' work and towards assuring the integrity of the processes for assessment and award'. For example, it is often no longer considered appropriate for external examiners to change individual student's marks (QAA 2012) even though institutional regulations generally permit this.

One might argue that a common understanding of the role in relation to standards is not necessary. The potential conflict between the general academic advising role and the ability to take an objective stance in relation to student performance was considered and dismissed as no 'real danger' by the Lindop Report (Lindop Committee 1985); indeed it is a challenge that academics face on a daily basis. However, there is a danger in this changing context that some interpretations may be less effective in terms of examiners' contribution to verifying and maintaining academic standards. Examiners who operate in a consultant or critical friend role are more likely to focus on the standards of academic 'input' and 'process' rather than 'output'. The failure of quality assurance, such as external examining, to focus on academic standards as output is a criticism posed by a number of writers (Harvey 2002; Alderman 2009; Middlehurst 1996). Alderman (2009) argues that such a focus is underpinned by the view that high standards will be maintained by standardised procedures, a view which he considers 'false and dangerous' (13). His opinion is supported by other research, for example on the extent to which double marking achieves consistency (Yorke 2008) or accuracy (Satchell and Pratt 2010) and on the failure of tightly managed assessment procedures to create an effective experience for students (Crook et al. 2006).

On the other hand, an approach that positions examiners as arbiters of standardised procedures rather than arbiters of standards may be an implicit acceptance that giving them such strong power

over determining standards is not working; a potential response to lack of trust in either the examining system or individual examiner's knowledge and judgement. This is reinforced by the new Quality Code establishing external examiners as 'only one part, albeit a very important part' of the quality assurance of the UK higher Education system (QAA 2012: section B7, 3). Whilst, the Code also sets out an expectation that higher education providers make scrupulous use of external examiners, it confirms a wide ranging view of examiners' functions, clearly stating that it does not 'assume a specific model of assessment oversight' (5). This situation, adopted in acknowledgement of institutional differences, challenges the assumption that there is either a specified role for examiners or a common understanding of the examiners' role.

Assumption 3

There are processes in place that support the development of consensus regarding standards which enables them to be understood and consistently applied by examiners.

Efforts to record and monitor standards began in the UK with the establishment of the Council for National Academic Awards (CNAA) in 1964 with subsequent work seeking to make academic standards more explicit and consistent in an effort to support consensus in the sector. In particular, developments include a Framework For Higher Education Qualifications at the system level, Subject Benchmark Statements and Professional standards at the discipline level and assessment criteria, grade descriptors and specified learning outcomes at the institutional, departmental and programme level. These developments have all been based on the belief that previously elusive standards could be made more explicit through techniques such as benchmarking (Woolf and Cooper 1999) although there was some reluctance from subject associations fearing that efforts to create greater comparability of standards in this way might lead to over standardisation of the curriculum (HEQC 1996b). The Silver Report (Silver et al. 1995) asserts that '[c]onsistency (in

standards) is the outcome of institutions defining and implementing their own standards in the context of appropriate external frameworks and scrutiny' (p42). In relation to external examining, such codification is considered important in guaranteeing consistency of standards (Universities UK and others 2010).

However such explicit standards promise more than they can deliver (Hawe 2002) and O'Donovan *et al.* (2004) stress the futility of trying to define standards precisely thereby confirming problems identified by Sadler (1987). He argues that (205) fuzzy standards cannot be transformed into sharp standards by simply using more detailed language.

Moss and Schutz (2001) provide a useful exploration of the creation and use of explicit standards for judgement, drawing on teaching standards as their exemplar. Their philosophical argument is based on Habermas' (1996) view that the moral justification for laws (or in this case standards) comes from the fact that they are created through rational consensus with relevant representatives of the community involved in drawing up the standards; 'a fair dialogue among equals' (Moss and Schutz, 40) . They argue that attempting to seek a consensus leads groups to produce too general or abstract standards, often based on compromise rather than consensus, which hide complexity and can mask diversity as those involved strive to complete the task in a limited time frame. Once standards are written, the system operates outside the control of the standard setting group. This means that we need standards which are created by this democratic dialogue but which can be used independently as 'fixed guidelines' (p42). Drawing on Habermas' presuppositions about discourse ethics, Moss and Schutz discuss the assumption underlying written standards that those involved 'confer (red) identical meanings on the expressions they employ(ed)' (Habermas 1996, 19) and that the 'expressions (will) keep the same meaning in the diverse situations in which they are employed' (p11). Standards have to be used interpretively for individual examples (as in judges interpreting the law in relation to specific cases) but this interpretation should not fall outside what was determined by the community consensus that drew them up. However, in their empirical study

of just such a group, they showed how it failed to produce something which can be used independently by assessors for several reasons. These include the fact that assessors' understanding of terms differs because of their previous experience, also found in Hawe (2002); although group members may agree to use the same words, this doesn't assure agreement about their meaning. These findings have been reflected in a range subjects and contexts (see for example Martin and Cloke 2000; Tang 2008; Brooks 2012). Overall Moss and Shutz (2001) argue that judgement becomes an 'interpretation' of standards not a 'reflection' of them.

Of further interest is the failure of explicit statements to be grounded in empirical study of the standards assessors actually use (Milanovic et al. 1996) or students' actual performance (Greatorex 2000) which is fundamentally important. As Sadler (1987, 196) says, standards must be grounded in field experience not in theory; that is, what people actually draw on to make their judgement, not what they think the standards should be. For example, Shay (2008, 598) argues that criteria don't reflect everything valued by academics (e.g. subject knowledge). She suggests the power of these documents to secure consistency may be limited by their focus on 'academic and intellectual skills', privileging those achievements at the expense of discursive or propositional knowledge.

In relation to research specifically on external examiners, there are interesting findings regarding their use of information to enable them to represent community standards. For example, Ross (2009) argues that examiners are still bounded by their social and cultural environment and expectations, and Colley and Silver's (2005) research identifies the importance of personal experience of both standards and quality assurance processes in providing examiners' reference points, with less significance given to *Academic Infrastructure* and other formal reference points (also see Hawe 2002; QAA 2005). The most important information, from the point of view of examiners was the assessment guidance and criteria for individual tasks within courses although it could be argued that this is more likely to represent local rather than community standards. Module and departmental information was seen as more important than institutional information and, in

many cases, professional body standards were not used by external examiners as they were seen as only loosely related to academic standards. At root, it seems that despite a couple of decades attempting to build an infrastructure of explicit standards, the successful operation of external examining still rests with the individuals concerned as suggested in 1996:

‘Ultimately, academic standards are established through the exercise of expert judgement by the examiners in their own field of expertise. This remains true regardless of whether the institution has sophisticated quality assurance procedures or whether explicit criteria have been evolved to aid the setting of standards. Such procedures and specifications should be viewed as a means of supporting and demonstrating high professional standards’ (HEQC 1996c, 72).

Examiners themselves

In the light of the last comment, it is important to consider certain existing assumptions regarding the capacities of individual examiners to meet the purposes of the examining system. This includes their knowledge of standards, their capacity to apply those standards appropriately and consistently and their understanding of the complexity of professional judgement and assessment procedures more generally.

Assumption 4

External Examiners have ‘extensive knowledge of academic standards, at least within a discipline or subject area, and across the sector’ (HEQC 1996a, 31)

This assumption underpins the idea that external examiners can provide consistent independent judgement which justifies privileging their perspective as ‘guardian of the ‘standards’’ (Shay 2005,

674). Norcini and Shea (1997) argue that in order to gain credibility for standards, standard setters should be suitably qualified. However the very nature of external examining, with a small number of examiners often covering a whole degree scheme, means that they may well be passing judgement on the standards of work outside their domain of expert knowledge. In addition, Colley and Silver (2005) point out that examiners, by and large, only examine in a limited range of universities with exposure only to standards in those institutions. Reports on external examining (Silver et al. 1995; QAA 2007) confirm these concerns regarding the power of examiners to secure community standards; examiners are working within a bounded sector of HE institutions and thus may not have wide enough experience of different institutions.

Indeed, efforts to improve their knowledge of standards may be restricted by notions of academic freedom. Warren Piper (1994) suggests that examiners actively resist the introduction of technical routines for regulating judgement as a threat to their autonomy. Similarly Howe (2002) argues that academics more generally may reject explicit standards as deskilling and a threat to their professional autonomy or humanistic views of teaching and learning. From this perspective, perhaps the ongoing deliberations about external examining are essentially part of a 'tussle' (Sadler 2011) between academic autonomy and institutional or government authority? Interestingly, there is an argument that academic freedom includes the freedom, as experts, to assign grades without interference by the wider university although the law in the UK and legal cases in the USA contradict this interpretation (Hill 2011).

The assumption of examiner knowledge relies on the notion that standards exist objectively and separately from individual examiners, albeit with possible differences across disciplines and universities. The foregoing debate about the individualised nature of standards would posit a more informal view of examiners' knowledge of standards. The interpretive perspective would tend to emphasise the notion of examiners learning, not through reference to documentation and explicit standards, but through being part of an assessment community. Indeed, Colley and Silver (2005)

found that examiners claim to learn from seeing other externals at work and some from learning and teaching networks. Induction was only mentioned by one of the groups of examiners researched. However, whilst Silver and Hannan (2004) found a range of motivations for becoming an examiner including making a contribution to the academic community and gaining intelligence about other institutions' programmes, none of them included calibrating their sense of standards. Therefore, as Ross (2009) argues, examiners bring their beliefs, values and assumptions to the task and it is presumptive to assume that extensive knowledge of academic standards is included.

Assumption 5

External examiners can represent community standards reliably and consistently and do not suffer from the same type of frailties exhibited by other assessors.

In order to explore this assumption, we are drawing on research on higher education grading because passing judgement on the quality of student work remains central to how examiners discharge their responsibilities.

Drawing on the hermeneutic approach discussed earlier, marking student work is conceived of as a socially situated interpretive act where the meaning of standards is constituted through the discursive practice (dialogue and language) which takes place in the social, cultural and political contexts concerned. Shay (2005, 664) states 'interpretations [of standards] are contextually constituted and cannot be divorced from the value-bases which interpreters bring with them. Differences between markers are not 'error', but rather the inescapable outcome of the multiplicity of perspectives that assessors bring with them'. For this reason, tutors' perspectives on standards differ depending on other social worlds that assessors inhabit (Den Outer et al. 2012) their history and values (Dobson 2008b) and their previous experience (Milanovic et al. 1996). Drawing on a different epistemological approach, Sadler (1987, 204) argues that people achieve their knowledge

of standards' through 'inductive cognitive processes', reinforcing the importance of personal experience in developing a sense of standards.

Thus, individuals' *standards' frameworks*, as defined earlier, are constructed and reconstructed over time by reference, amongst other things, to specialist knowledge, other tutors' marking, students' work, involvement in moderation and examiners' feedback. This creates a unique lens through which tutors and thus examiners, read and judge student work (Bloxham and Boyd 2011). Undoubtedly, this personalised knowledge of standards is mediated by the power of shared, situated learning. Eisner (1985) refers to this process as the use of 'connoisseurship'; the well-informed subjective judgement which accrues through immersion in a subject discipline. A number of writers have tackled this issue in the context of higher education assessment (Shay 2005; Orr 2010). The argument suggests that whilst teachers develop their own standards frameworks, 'individual lecturers' assessment responses are constituted collectively' (Orr 2010, 15) as activities such as moderation and external examining help to build an intersubjective consensus. Thus 'standards' frameworks merge as common standards are continually negotiated (Orr 2010). For example, Dobson (2008a, 337) provides examples of how assessors working together over a long period had established 'rarely articulated in-the-head normative agreement' about grades. The question is whether this in-the-head agreement reflects a merging of standards which is sufficient or widespread enough in general to claim that markers are representing wider community standards. Furthermore, in relation to external examiners, the question is whether as members of an academic community, they have engaged sufficiently with the practices of that community to acquire a strong and shared understanding of standards which they can transfer to, and articulate in, new settings. The evidence for 'intersubjectivity' (Bruner 1996, 20) is not very compelling in the context of higher education assessment. It is an untenably positive view of how shared standards are acquired and used in practice, whether by academics in general or by external examiners in particular. Whilst the

material in the following paragraphs is not propounded as a criticism of academic teachers but a recognition of the complex and intuitive nature of judgement at this level, it does indicate a range of ways in which judgement in grading is shown to be problematic. How confident can we be that external examiners are not subject to the same vagaries in practice?

Overall, assessors' unreliability is well documented (Brooks 2012). Milanovic *et al.* (1996) point to the marker as a recognised source of error and Moss and Shultz (2001) indicate that, even where assessors agree pass and fail, they do so for different reasons. Milanovic *et al.* (1996) identify a range of studies showing that markers focus on different aspects of student composition, for example, content, length, accuracy, vocabulary or no one thing but more the 'environment of the essays – the effect of the papers' (Vaughan 1991). Elander and Hardman (2002) also report that tutors vary in what they look for (cues) in assessing work and this contributes to variance in marks.

Furthermore, research points to the influence of students' work on assessors' judgement. The 'techno-rationalist paradigm' as expressed in most guidance to tutors advocates that we should base our assessment on criterion rather than norm referencing so that a student is judged against a set of standards, not against his or her peers. This distinction has been criticised (Orr 2008; Neil and Wadley 1999; Yorke 2009). In particular, Yorke makes the point that 'assessors' grading behaviour is tacitly influenced by norm referencing and Vaughan (1991), Shay (2004), Delaney (2005) and Orrell (2008) also found that tutors draw on their knowledge of different students' work in order to make their judgements. More broadly, Crisp (2008) and Bloxham *et al.* (2011) found that other students' work affects individuals' standards' frameworks. There is no evidence that external examiners will not be influenced in the same way in making their judgements. In particular, following this argument, examiners' (sometimes limited) exposure to a range of student work as comparators is likely to affect their sense of standards.

A further weakness in academics' ability to represent community standards relates to their limited use of codified standards within and across institutions. Empirical studies (see Hawe 2002; Broad 2003; Greatorex 2000; Dobson 2008a) have found that assessors use personal criteria beyond or different to those stated including contextual factors. Colley and Silver's (2005) evidence that external examiners draw mostly on their own experience rather than documented statements of standards tends to reinforce this point and contradicts the assumption that they do not suffer from the same type of frailties exhibited by other assessors.

Some studies (see Dobson 2008a; Broad 2003; Elander and Hardman 2002) have shown that knowledge of the student influences tutors' judgement. As 'outsiders', external examiners might be considered less susceptible to this weakness although the degree to which they resist the appeals of the home assessors in relation to taking into account aspects of the student has yet to be researched except in relation to viva voce examinations (Dobson 2008a). He found that where there is a face to face encounter, examiners cannot avoid taking into account aspects of individuals, suggesting that contextual factors are more likely to influence examiners in a viva although they will have impact in other judgements. Dobson (2008a) shows how grades are negotiated in dialogue and that norm referencing remains prevalent in examiners' decisions (Dobson 2008b).

More generally, psychological enquiry tends to challenge assumptions of expert judgement considered to underpin external examining. Using the 'Heuristics and Biases' approach, research (Gilovich et al. 2002) suggests that expert judgement in cases of uncertainty or complexity utilises a number of simplifying cues (heuristics) because a more rational and detailed judgement is too difficult or time consuming. Such heuristics allow for systematic biases to creep in of which the expert is unaware and may lead to assessors predicting performance in advance of completing reading (Crisp 2008). For example, recent research has identified that knowledge of the student, group stereotyping and surface features of work can be sources of bias (Brooks 2012; Kangis 2001; Hartley et al. 2006).

The extent to which external examiners do not suffer from the same type of frailties exhibited by other assessors is under-researched but examiners' reliance on personal experience (Colley and Silver 2005) tends to confirm the importance of personal standards frameworks as compared with shared standards. Yet, despite this lack of evidence that they can represent disciplinary community standards, as opposed to local or personal, standards, we ask them to play an elevated role with an expectation that they are able to hold and use standards in a reliable and mutually consistent way.

Assumption 6

External examiners understand the complexity of professional judgement and higher education assessment processes.

This discussion of the assumptions underpinning external examining has uncovered the complex nature of academic standards and its implications. External examiners hold major responsibilities; they are held in considerable esteem by tutors (Bloxham and Boyd 2011) and by the sector in general (Watson 2006; Gaunt 1999) and their existence is considered to be a key quality assurance mechanism. Yet what confidence can we have that the average external examiner has the 'assessment literacy' to be aware of the complex influences on their standards and judgement processes? How likely are they to understand the provenance of their own standards and the influence of their background and experience? How strong is the temptation to draw largely on their personal experience as an indicator of what standards should be rather than recognising the potential bias in that approach, or the influence of a particular context or student body or professional experience? Sadler (2011) discusses the rationale for academics' ability in general to act autonomously in judging the standards of student work and it includes their qualifications, research and exposure to standards elsewhere in the sector. He challenges this rationale with arguments that academics are not qualified to act autonomously because they are not qualified or

skilled in assessment. Likewise, less recently, Yorke et al. (2000) suggest that few academics have expertise in assessment and in the first major, detailed survey of the external examining system (ESRC 1989), it was reported that 'the judgement and actions of examiners are largely informed by experience and knowledge of their subject and very little informed by the equally pertinent body of knowledge about the examination and the measurement of human performance' (p30).

Colley and Silver (2005) also argue for the need for external examiners to accept other assessment cultures and not to carry too much 'baggage'. Indeed Biggs (2002) writes convincingly about the potential of examiners to counteract enhancement in assessment practice because of their lack of knowledge of assessment. The Dearing report (NCIHE 1997) claimed that standards are a shared responsibility of the whole HE community but such a view doesn't deal with the complexity of problems created by knowing, using and continually reconstructing standards as external examiners. Unfortunately successive enquiries into examining have not tackled these issues.

Conclusion

The external examiner system plays a key role in assuring academic standards. This article has identified and questioned the validity of the assumptions which underpin the system. Essentially, external examining assumes that there is a consensus on standards and that we can vest in individuals the ability to represent that consensus in judging the comparability of academic standards in a stable and appropriate way for any given programme or set of modules. Yet the critique of these six assumptions challenges the notion of the examiner as an independent objective voice (Shay 2005) and one might argue that such a challenge holds greater sway now than it did in the past given the massification of higher education and the limited dialogue and sharing of standards across universities. As Shay (2005, 674) says, quality 'systems rely on consistency between independent judgement and privilege (at least officially) the perspective of the external

examiner who is appointed guardian of the 'standards'. Yet the assumptions on which that guardianship is founded are, if not disproven, certainly lacking in sound supporting evidence.

It is difficult to argue against the external examiner system at the level of principle. Given all that has been discussed in this paper about the social construction of standards and, therefore, the important benefits of inter-institutional sharing and debating of them as the basis for a community consensus, the systematised opportunity for academics to examine their peers' standards makes good sense. It is in the detail that this principle is threatened. Inquiries into the role (for example, Universities UK 2011) have chosen to tackle problems through tightening up procedures and induction and making the role and examiners' reports more transparent and accessible. They have not focused on addressing the fundamental assumptions on which the principled role is based and therefore, we can have little confidence that enhanced procedures will deliver the desired change. The system is expensive and, if the resources are to continue to be devoted to it, it might be sensible to research further the conditions under which external examining could meet its aims for institutions and for assuring a broad comparability of standards more widely.

References

- Alderman, G. 2009. Defining and measuring academic standards: a British Perspective. *Higher Education Management and Policy* 21, no. 3: 11-22.
- Ashworth, M., Bloxham, S. and Pearce, L. 2010. Examining the tension between academic standards and inclusion for disabled students: the impact on marking of individual academics' frameworks for assessment. *Studies in Higher Education* 35, no. 2: 209-223.
- Attwood, R. 2010. Nip and tuck for externals? A radical facelift is required. *Times Higher Educational Supplement*, 12th August.

Biggs, J. 2002. *The reflective institution: assuring and enhancing the quality of teaching and learning*.

Available at:

http://www.heacademy.ac.uk/resources.asp?process=full_record&id=109

BIS. 2009. *Higher Ambitions: The future of universities in a knowledge economy*. London: Department for Business Innovation and Skills.

Bloxham, S. 2009. Marking and moderation in the UK: false assumptions and wasted resources.

Assessment & Evaluation in Higher Education 34, no. 2: 209-220.

Bloxham, S. and Boyd, P. 2011. Accountability in grading student work: securing academic

standards in a twenty-first century quality assurance context. *British Educational Research*

Journal, i-first, DOI:10.1080/01411926.2011.569007.

Bloxham, S., Boyd, P. and Orr, S. 2011. Mark my words: the role of assessment criteria in UK higher

education grading practices. *Studies in Higher Education* 36, no. 6: 655-670.

Broad, B. 2003. *What we really value: Beyond rubrics in teaching and assessing writing*. Logan, Utah:

Utah State University Press.

Broad, B. 2000. Pulling your hair out: Crises of standardization in communal writing. *Research in the*

Teaching of English 35: 213-260.

Brooks, V. 2012. Marking as judgement. *Research Papers in Education* 27, no. 1: 63-80.

Bruner, J.S. 1996. *The Culture of Education*. Cambridge, MA: Harvard University Press.

Coates, H. 2010. Defining and monitoring academic standards in Australian higher education. *Higher*

Education Management and Policy 22, no. 1: 41-58.

Colley, H. and Silver, H. 2005. *External examiners and the benchmarking of standards*. York: Higher

Education Academy.

Crisp, V. 2008. Exploring the nature of examiner thinking during the process of examination

marking. *Cambridge Journal of Education* 38, no. 2: 247-264.

- Crook, C., Gross, H. and Dymott, R. 2006. Assessment Relationships in Higher Education: The Tension of Process and Practice. *British Educational Research Journal* 32, no. 1: 95-114.
- Delandshere, G. 2001. Implicit theories, unexamined assumptions and the status quo of educational assessment. *Assessment in Education* 8, no. 2: 113-133.
- Delaney, C.M. 2005. The evaluation of university students' written work. *British Education Association Conference*. University of Glamorgan 14-17 Sept.
- Den Outer, B., Handley, K. and Price, M. 2012. Situational analysis and mapping for use in education research: a reflexive methodology? *Studies in Higher Education*, available on iFirst, ISSN 0307-5079.
- Dobson, S. 2008a. Applying a validity argument model to three examples of the viva. *Nordisk Pedagogik* 28: 332-344.
- Dobson, S. 2008b. Theorising the viva - A qualitative approach. *Assessment & Evaluation in Higher Education* 33, no. 3: 277-288.
- Ecclestone, K. 2001. I know a 2:1 when I see it: understanding criteria for degree classifications in franchised university programmes. *Journal of Further and Higher Education* 25, no. 3: 301-313.
- Eisner, E.W. 1985. *The art of educational evaluation: a personal view*. London: Falmer.
- Elander, J. and Hardman, D. 2002. An application of judgement analysis to examination marking in psychology. *British Journal of Psychology* 93: 303-328.
- ESRC. 1989. *The role of External Examiners*. Swindon: Economic and Social Research Council.
- Gaunt, D. 1999. The practitioner as external examiner. *Quality in Higher Education* 5, no. 1: 81-90.
- Gilovich, T., Griffin, D. and Kahneman, D. (ed.) 2002. *Heuristics and biases*. Cambridge: Cambridge University Press.
- Gipps, C. 1999. Socio-cultural aspects of assessment. *Review of Research in Education* 24: 355-392.

- Greatorex, J. 2000. *Is the glass half full or half empty? What examiners really think of candidates' achievement*. Cardiff September 7-10.
- Habermas, J. 1996. *Between facts and norms: contributions to a discourse theory of law and democracy* (W. rehg, Trans). Cambridge: MIT Press (original work published 1992).
- Hannan, A. and Silver, H. 2004. *Enquiry into the nature of external examining: final report*. York: Higher Education Academy.
http://www.heacademy.ac.uk/resources.asp?process=full_record&id=376
- Hartley, J., Trueman, M., Betts, L. and Brodie, L. 2006 What Price Presentation? The effects of typographic variables on essay grades. *Assessment & Evaluation in Higher Education* 31, no. 5: 523-534.
- Harvey, L. 2002. Evaluation for what? *Teaching in Higher Education* 7, no. 3: 246-263.
- Harvey, L. 2004–12. Analytic Quality Glossary. Quality Research International.
<http://www.qualityresearchinternational.com/glossary/> (accessed January 22, 2013)
- Hawe, E. 2002. Assessment in a pre-service teacher education programme: the rhetoric and the practice of standards-based assessment. *Asia Pacific Journal of Teacher Education* 30: 93-106.
- HEFCE. 2009. *Report of the sub-committee for Teaching, Quality, and the Student Experience: HEFCE's statutory responsibility for quality assurance*. Bristol: HEFCE (2009/40).
- HEQC. 1996a. *Strengthening external examining*. London: Higher Education Quality Council.
- HEQC. 1996b. *Threshold and other academic standards: The views of four subject groups*. London: Higher Education Quality Council (DQE226 3/96).
- HEQC. 1996c. *Understanding academic standards in modular frameworks*. London: Higher Education Quality Council.
- HEQC. 1994. *Learning from audit*. London: Higher Education Quality Council.

- Hill, D. 2011. A Contentious Triangle: Grading and Academic Freedom in the Academy. *Higher Education Quarterly* 65, no. 1: 3-11.
- Innovation, Universities, Science and Skills Committee. (2009) *Students and Universities*. London: House of Commons.
- Kangis, P. 2001. Presentational Dimensions and Marks Awarded to Assignments. *Quality in Higher Education* 7, no. 3: 199-206.
- Lindop Committee. 1985. *Academic Validation in public sector higher education (Lindop Report)*. London: HMSO, (Cmnd 9501).
- Martin, S. and Cloke, C. 2000. Standards for the award of qualified teacher status: reflections on assessment implications. *Assessment & Evaluation in Higher Education* 25, no. 2: 183-190.
- Messick, S. 1989. Validity. In *Educational Measurement*, ed. R. Linn, New York: Macmillan.
- Middlehurst, R. 1996. Degree standards and quality assurance: a discussion. In *Changing Conceptions of Academic Standards*, ed. J. Brennan, 71-96. London: Quality Support Centre, Open University.
- Milanovic, M., Saville, N. and Shuhong, S. 1996. A study of the decision-making behaviour of composition markers. In *Performance Testing, Cognition and assessment*, ed M. Milanovic, and N. Saville, 92-114. Cambridge: Cambridge University Press.
- Moss, P.A. and Schutz, A. 2001. Educational Standards, Assessment and the search for consensus. *American Educational Research Journal* 38, no. 1: 37-70.
- National Union of Students. 2009. *External examiners: their role in UK higher education*. London: NUS, (QMS/09/02).
- NCIHE. 1997. *Higher Education in the Learning Society: Report of the National Committee of enquiry into higher education (Dearing Report)*. Leeds: NCIHE Publications.
- Neil, D.T. and Wadley, D.A. 1999. A generic framework for criterion-referenced assessment of undergraduate essays. *Journal of Geography in Higher Education* 23, no. 3: 303-319.

- Norcini, J.J. and Shea, J.A. 1997. The credibility and comparability of standards. *Applied Measurement in Education* 10, no. 1: 39-59.
- O'Donovan, B., Price, M. and Rust, C. 2004. Know what I mean? Enhancing student understanding of assessment standards and criteria. *Teaching in Higher Education* 9, no. 3: 325-335.
- Orr, S. 2010. "We kind of try to merge our own experience with the objectivity of the criteria": The role of connoisseurship and tacit practice in undergraduate fine art assessment. *Art, Design & Communication in Higher Education* 9, no. 1: 5-19.
- Orr, S. 2008. Real or imagined? The shift from norm referencing to criterion referencing in higher education. In Havnes, A. and McDowell, L. *Balancing Dilemmas in Assessment and Learning in Contemporary Education*, 133-144. London: Routledge.
- Orrell, J. 2008. Assessment beyond belief: The cognitive process of grading. In *Balancing dilemmas in assessment and learning in contemporary education*, ed A. Havnes and L. McDowell, 251-263. London: Routledge.
- QAA. 2012. *UK Quality Code for Higher Education*. Available at:
<http://www.qaa.ac.uk/AssuringStandardsAndQuality/quality-code/Pages/default.aspx>
- QAA. 2009. *Thematic enquiries into concerns about academic standards in higher education in England*. Gloucester: QAA.
- QAA. 2007. *The Classification of Degree Awards*. Gloucester: Quality Assurance Agency.
- QAA. 2005. *Outcomes from institutional audits: external examiners and their reports*. Gloucester: Quality Assurance Agency.
- Ross, V. 2009. External music examiners: micro-macro tasks in quality assurance practices. *Music Education Research* 11, no. 4: 473-484.
- Sadler, D.R. 2011. Academic freedom, achievement standards and professional identity. *Quality in Higher Education* 17, no. 1: 85-100.

- Sadler, D.R. 1989. Formative assessment and the design of instructional systems. *Instructional Science* 18, no. 2: 119-144.
- Sadler, D.R. 1987. Specifying and promulgating achievement standards. *Oxford Review of Education* 13, no. 2: 191-209.
- Satchell, S. and Pratt, J. 2010. The dubiety of double marking. *Higher Education Review* 42, no. 2: 59-62.
- Shay, S.B. 2004. The Assessment of Complex Performance: A Socially Situated Interpretive Act. *Harvard Educational Review* 74, no. 3: 307-329.
- Shay, S. 2008. Beyond social constructivist perspectives on assessment: the centring of knowledge. *Teaching in Higher Education* 13, no. 5: 595-605.
- Shay, S. 2005. The Assessment of Complex Tasks: A Double Reading. *Studies in Higher Education* 30, no. 6: 663-679.
- Silver, H., Stennett, R. and Williams, R. 1995. *The external examiner system: possible futures (The Silver Report)*. London: Higher Education Quality Council.
- Silver, H. and Williams, S. 1996. Academic Standards and the external examiner system. In *Changing conceptions of academic standards*, ed J. Brennan, et.al., 27-48. London: Quality Support Centre, Open University.
- Tang, S.Y.F. 2008. Issues in field experience assessment in teacher education in a standards based context. *Journal of Education for Teaching* 34, no. 1: 17-32.
- Universities UK. 2011. *Review of External examiner arrangements in universities and colleges in the UK: Final Report and Recommendations*. London: Universities UK.
- Universities UK, Guild of higher Education, QAA. 2010. *Review of External Examining Arrangements in the UK: A discussion paper from Universities UK, Guild HE and the Quality Assurance Agency for Higher Education*. London: Universities UK.

- Vaughan, C. 1991. Holistic assessment: what goes on in the rater's mind? In *Assessing second language writing in academic contexts*, ed L. Hamp-Lyons, 111-125. Norwood, N.J.: Ablex Publishing Corp.
- Warren Piper, D. 1994. *Are professors professional*. London: Jessica Kingsley.
- Watson, D. 2006. *Who killed what in the quality wars*. Gloucester: Quality Assurance Agency.
- Wolf, A. 1995. *Competence-based assessment*. Buckingham: Open University Press.
- Woolf, H. and Cooper, A. 1999. Benchmarking academic standards in history: an empirical exercise. *Quality in Higher Education* 5, no. 2: 145-154.
- Yorke, M. 2009. Faulty signals? Inadequacies of grading systems and a possible response. In *Assessment, learning and judgement in higher education*, ed G. Joughin, 65-84. Netherlands: Springer.
- Yorke, M. 2008. *Grading Student Achievement in Higher Education*. London: Routledge.
- Yorke, M., Bridges, P. and Woolf, H. 2000. Mark distributions and marking practices in UK higher education. *Active Learning in Higher Education* 1, no. 1: 7-27.